PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/10, C12Q 1/68

A1

(11) International Publication Number: **WO 98/05764**

(43) International Publication Date: 12 February 1998 (12.02.98)

(21) International Application Number: PCT/DK97/00316

(22) International Filing Date: 23 July 1997 (23.07.97)

(30) Priority Data:
8/208423 7 August 1996 (07.08.96) JP

(71) Applicant *(for all designated States except US)*: NOVO NORDISK A/S [DK/DK]; Novo Allé, DK-2880 Bagsvaerd (DK).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MIYOTA, Yoshiaki [JP/JP]; Showa Denko K.K., Central Research Laboratory, 1-1, Ohnodai 1-chome, Midori-ku, Chiba-shi, Chiba 267 (JP). FUKUYAMA, Shiro [JP/JP]; Showa Denko K.K., Central Research Laboratory, 1-1, Ohnodai 1-chome, Midori-ku, Chiba-shi, Chiba 267 (JP).

(74) Common Representative: NOVO NORDISK A/S; Corporate Patents, Novo Allé, DK-2880 Bagsvaerd (DK).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: NUCLEIC ACID POOL AND METHOD FOR PRODUCING THE SAME

(57) Abstract

To efficiently provide a number of base sequences existing in spaces significantly apart from naturally-existing base sequence spaces. A nucleic acid pool comprising two or more different nucleic acids, which is constructed by dividing all or a part of one or more genes into 3 or more blocks followed by ligating all or a part of these blocks into sequences that are different from the sequence or sequences of the original, non-divided gene or genes, and a method for producing the pool; and also a genetic product to be obtained by expressing the genetic information existing in the nucleic acid pool.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Title: Nucleic acid pool and mehtod for producing the same

INDUSTRIAL FIELD

The present invention relates to a nucleic acid pool to be produced by ligating oligonucleotides at random, and a method for producing it, and also to a genetic product to be produced by expressing the nucleic acid existing in the pool as a gene..

BACKGROUND ART

One approach to protein engineering for improving naturally-existing proteins to modified ones which are more useful to human beings is to improve proteins through site-specific mutation, which has produced some results (Japanese Patent Application Laid-Open No. 5-91876). However, this requires the clarification or identification of the stereostructure of the targeted protein, and much labor is needed for the analysis of the stereostructure. In addition, even though the stereostructure could be clarified or identified, there are still many unknown matters for the relationship between the structure and the function with proteins. Therefore, it is still difficult to surely impart an intended function to the targeted protein.

In order to overcome these difficulties, a process comprising random mutation and screening and also evolutional molecular engineering that utilizes the evolution of organisms have been being highlighted and said to be extremely useful (Proc. Natl. Acad. Sci., USA, 83, 576 (1986)). However, the

current methods are directed to the substitution of at most several amino acids.

In WO95/22625, disclosed is a method for forming novel genes by dividing a plurality of genes at random and homologously recombining them to reconstruct novel genes. However, this is one method for forming chimera genes. The genes to be formed by this method are similar to the original genes, and the former shall have the essential base sequences of the latter.

Using such known methods, it is difficult to desire the impartation of some additional functions to organisms which they could not gain during the steps of their evolution. In order to obtain genetic products, of which the functions are greatly different from those of naturally-existing substances such as proteins, it is believed effective to prepare a pool of nucleic acids having significantly different base sequence spaces from those existing naturally, and to produce from them genetic products having the intended functions.

One method for this may be to prepare a nucleic acid pool that covers all base combinations. However, even the total number of the base sequences that may code for a relatively small protein with 100 amino acids (300 bp) is an enormous number of 4^{300} (about 10^{180}), and it is in fact impossible to prepare the nucleic acid pool that may cover all of them.

For proteins of some kinds, their sub-structures which are referred to as modules were specifically noted, and an attempt was made to change the sequencing of the base sequence blocks corresponding to the individual modules to thereby produce

mutants having different module sequences (Viva Origino, Vol. 23, No. 1 (1995) 86-87). In this attempt, however, the base sequences were re-sequenced merely individually for the individual mutants. No one has heretofore attempted the formation of a nucleic acid pool covering all re-sequenced molecules and the collection of genes capable of expressing products having intended properties from the pool.

The subject matter of the present invention is to provide a method for efficiently obtaining base sequences that exist in spaces greatly different from those of naturally-existing base sequences, and also to provide genetic products to be obtained by expressing, as genes, the nucleic acid sequences that are obtained in that manner and that do not exist naturally.

The sequence space of a gene includes the full-length sequence thereof to be theoretically constituted by a combination of four bases, A, G, C and T. For example, a base sequence that codes for a protein composed of a number "n" of amino acids shall be constructed by selecting and sequencing any desired one of the four bases for a total of 3n-times, therefore including 4^{3n} combinations. Accordingly, a protein composed of 100 amino acids shall include different base sequences of about 10^{18} types as so mentioned hereinabove.

In fact, there is no limitation for the number of amino acids that constitute proteins. Therefore, the sequencing spaces for proteins shall extend unlimitedly. During the steps of evolution of organisms, only a part of such sequencing spaces have been examined, and there is a great probability that some sequences coding for proteins which may have some

extremely excellent functions could exist in the other great sequencing spaces. The protein engineering studies which have been and are being made in many laboratories and institutes at present are essentially directed to the creation of novel proteins having functions superior to those of naturally-existing proteins, and one essential approach made therein to this purpose is to substitute amino acids in existing sequences, as so mentioned hereinabove. However, the amino acid substitution is nothing but the essential means that organisms have carried out during the steps of their evolution or, that is, such is the imitation of organisms and is to search only around the sequences that organisms already examined. In addition, there is a probability that the sequences thus obtained will be those that were already weeded out in the past.

We, the present inventors have considered that, in order to be greatly apart from the sequencing spaces that organisms already examined, if we carry out such matters that could not have been carried out by organisms, the purpose will be attained. We know that the division of a gene into several blocks followed by the change in the sequencing of the thus-divided blocks, if occurred in organisms, shall kill the organisms. Therefore, we have concluded that this method is suitable for our purpose. Having thus concluded, we, the present inventors have assiduously studied various matters relating to this method and, as a result, have succeeded in the finding of base sequences which are significantly apart from naturally-existing base sequencing spaces and also in the

formation of a molecule pool that covers such base sequences, and thus have completed the present invention.

Accordingly, the present invention provides a specific nucleic acid pool that is mentioned below, a method for producing it to be mentioned below, and also a genetic product to be obtained by expressing, as a gene, the nucleic acid existing in the nucleic acid pool, as is mentioned below.

1) A nucleic acid pool comprising two or more different nucleic acid, which is constructed by dividing all or a part of one or more genes into 3 or more blocks followed by ligating all or a part of these blocks into sequences that are different from the sequence or sequences of the original, non-divided gene or genes.

2) The nucleic acid pool according to the previous 1), which contains 10 or more different nucleic acids.

3) The nucleic acid pool according to the previous 1) or 2), which contains all the nucleic acids with different sequences as constructed by re-sequencing a number, n, of said blocks (where n is the number of the different blocks as formed by the division).

4) The nucleic acid pool according to any one of the previous 1) to 3), wherein the gene is a gene coding for a protein, and the amino acid sequence as encoded by each block is the same as the amino acid sequence as encoded by the corresponding part on the original gene.

5) The nucleic acid pool according to any one of the previous 1) to 4), wherein the gene is a gene coding for an enzymatic function or a control gene for it.

6) The nucleic acid pool according to the previous 5), wherein the gene is a gene coding for any one of proteases, lipases, cellulases, amylases, catalases, xylanases, oxidases, dehydrogenases, oxygenases and reductases.

7) The nucleic acid pool according to any one of the previous 1) to 6), wherein the gene is one derived from prokaryotes.

8) The nucleic acid pool according to the previous 7), wherein the gene is one derived from bacillus bacteria.

9) The nucleic acid pool according to the previous 8), wherein the gene is a protease API21 gene.

10) The nucleic acid pool according to any one of the previous 1) to 9), wherein each block is an oligonucleotide.

11) The nucleic acid pool according to the previous 10), wherein the nucleic acid is a single-stranded polynucleotide.

12) The nucleic acid pool according to the previous 10), wherein the nucleic acid is a double-stranded polynucleotide.

13) A method for producing a nucleic acid pool comprising two or more different nucleic acids, which comprises dividing all or a part of one or more genes into three or more oligonucleotide blocks or synthesizing oligonucleotides corresponding to said blocks, followed by ligating all or a part of these blocks into sequences that are different from those on genes.

14) The method for producing a nucleic acid pool according to the previous 13), which comprises the following steps a) to c):

a) a step of preparing 3 or more blocks of single-stranded oligonucleotides having base sequences that correspond to all or a part of one or more genes through division of one or more genes or through synthesis of oligonucleotide chains having said base sequences;

b) a step of adding a ribonucleotide to its 3'-terminal of each with a deoxyribonucleotide at the 3'-terminal of the oligonucleotide chain blocks as obtained in the previous step a), while adding a phosphoryl group to its 5'-terminal of each thereof with a hydroxyl group at the 5'-terminal; and

c) a step of ligating in any desired sequence the oligonucleotide chain blocks as obtained in the previous step b), by reacting the 3'-terminal ribonucleotide of one block with the 5'-terminal phosphoryl group of another block.

15) The method for producing a nucleic acid pool acording to the previous 14), wherein the number of the blocks to be prepared in the step a) is 3 or more.

16) The method for producing a nucleic acid pool according to the previous 14) or 15), wherein at least one block is left to still have its 5'-terminal hydroxyl group in the step b) to thereby selectively obtain a nucleic acid or nucleic acids having said block at the 5'-terminal.

17) The method for producing a nucleic acid pool according to any one of the previous 14) to 16), wherein at least one block is left to still have its 3'-terminal deoxyribonucleotide in the step b) to thereby selectively obtain a nucleic acid or nucleic acids having said block at the 3'-terminal.

18) The method for producing a nucleic acid pool according to any one of the previous 14) to 17), wherein the blocks are prepared in such a manner that the amino acid sequence as encoded by each block is the same as the amino acid sequence as encoded by the corresponding part on the original gene.

19) The method for producing a nucleic acid pool according to the previous 18), wherein blocks each having a base sequence of from the (3p+1)th to the (3q+2)th, as counted from the starting point of the reading frame on a gene (where p and q are integers to be independently determined for each block, provided that p, q), are prepared in the step a), and a ribonucleotide that corresponds to the (3q+3)th base, as counted in the same manner as above, or corresponds to a base that does not change the amino acid to be encoded at said site is added to each said block at its 3'-terminal.

20) The method for producing a nucleic acid pool according to any one of the previous 14) to 19), wherein the ligation of the step c) is conducted, using an RNA ligase in the presence of adenosine triphosphate and divalent metal ions.

21) A method for producing a double-stranded nucleic acid pool, which comprises converting the single-stranded nucleic acids as obtained in any one of the previous 14) to 20 into double-stranded ones through polymerase reaction.

22) A genetic product to be obtained by expressing the genetic information that exists in the nucleic acid pool of any one of the previous 1 to 12).

Now, the present invention is described in more detail hereinunder.

Nucleic Acid Pool

The "nucleic acid pool" as referred to herein means a high-density mixture of two or more different nucleic acids. Nucleic acids are single-stranded or double-stranded polyucleotides. The nucleic acid pool of the present inention can cover a specific number or more, for example, 10 or more different nucleic acid molecules having different structures. It is desirable that, when the mixture, nucleic acid pool is directly used in biochemical operation or reaction, it is in such a form that all the plural nucleic acid components constituting it can be reacted. However, the form of the mixture, nucleic acid pool is not specifically defined, and the nucleic acid pool may be either in solution or dry mixture.

The nucleic acid pool of the present invention is characterized in that it is constructed by dividing all or a part of one or more genes into 3 or more blocks followed by ligating all or a part of these blocks into sequences that are different from the sequence or sequences of the original, non-divided gene or genes, and is therefore characterized in that it comprises a plurality of different nucleic acids having base sequences that are different from the original, non-divided base sequence or sequences. The step of "ligating all or a part of the divided blocks into sequences that are different from the original, non-divided sequence or sequences" as referred to herein includes (i) re-sequencing of the blocks in

a sequence that is different from the original, non-divided sequence, (ii) ligation of a plurality of the same blocks continuously or discontinuously, (iii) re-ligation of the blocks except at least one block, and (iv) combination of these (i) to (iii). The operation for dividing a gene into plural blocks and re-sequencing these in any desired order that is employed in the present invention is hereinafter referred to as "shuffling".

For example, where one DNA has a sequence composed of a number, n, of blocks, as represented by a formula (1):

$$A - a1 - a2 - . . . . - a_n - B \qquad (9)$$

wherein the starting end A and/or the terminal end B may be omitted,

this may be shuffled according to the invention to give a mixture of nucleic acids to be represented by a formula (2):

$$A - a1' - a2' - . . . . - a_X - B \qquad (2)$$

wherein a1', a2', . . . , a_X are blocks that are independently selected from the group of a1, a2, . . . , a_n; and the total number of the blocks a1', a2', . . . , a_X may not be the same as the total number of the blocks a1, a2, . . . , a_n.

In order to make the shuffling effective, one or more genes must be divided into 3 or more blocks. If divided into 2 blocks, only one re-sequenced form can be obtained and many different nucleic acids cannot be obtained. If so, the effectiveness of the nucleic acid pool of the invention is poor.

Preferably, one or more genes are divided into 5 or more blocks.

The blocks (such as a1, a2, ... , a_n in the above-mentioned formula (1)), which are the units to be shuffled, are oligonucleotides or polynucleotides composed of 2 or more nucleotides (hereinafter referred to as "oligonucleotides"). If the length of each block is too short, the operation with the blocks is complicated. In general, therefore, each block is preferably composed of 21 or more nucleotide units, more preferably 45 or more nucleotide units. The uppermost limit of the block length is not specifically defined, provided that the block length is shorter than the length of one gene. If, however, the block length is too large, the re-sequenced nucleic acids to be obtained shall have many non-mutated base sequence parts. Therefore, in general, the block length is preferably within the range of from 5 to 30 % of the length of a gene.

The division of a gene into blocks may be effected at any sites of the gene. Though not excluding the division of a gene into the constitutive exons or segment blocks that correspond to the domains or modules of the protein which the gene codes for, there is a probability that the shuffling at such sites would have been examined in the natural world in the past. In order to obtain base sequences that have not heretofore been examined in the natural world, it is desirable that the division of a gene is effected inside the constitutive exons or at the sites corresponding to the inside of the domains or modules of the protein which the gene codes for.

During the shuffling of a gene, especially during the ligation of the divided blocks thereof, it is possible to introduce any oligonucleotide blocks which the original gene does not have, and also to insert or delete nucleotides. Needless-to-say, where the gene to be shuffled is a gene that codes for a protein, it is desirable that the gene blocks, oligonucleotides each have the same reading frame before and after the division of the gene. Namely, it is desirable that the gene blocks to be shuffled are so designed that they are translated to always give the corresponding amino acid sequences, irrespective of their relative positions in the shuffled sequence. Employing such means, it is possible to obtain proteins which have different structures as a whole from those of natural proteins but which partly contain amino acid sequences that have been confirmed to be useful in the natural world. Accordingly, the probability of obtaining useful proteins by such means is enlarged, as compared with the means of synthesizing proteins totally at random.

The re-sequencing of the divided blocks to be conducted through the shuffling thereof in the present invention is to ligate a desired number of the blocks, a1, a2, . . . , an, while allowing the ligation of two or more same blocks in series and allowing the deletion of some blocks, as so mentioned hereinabove. It is desirable that the nucleic acid pool of the invention to be obtained by the ligation covers at least all nucleic acids each composed of nearly the same number of blocks as the number of the divided blocks. For example, when a gene is divided into 5 blocks, a1, a2, a3, a4 and a5, it

is desirable that the nucleic acid pool obtained covers substantially all different combinations each comprised of these 5 blocks.

Precisely, it is desirable that the nucleic acid pool obtained covers all simple re-sequences, such as a1-a3-a2-a4-a5 (where the order of a2 and a3 was altered) and a1-a4-a2-a3-a5 (where s2, a3 and a4 were re-sequenced), more preferably complex re-sequences comprising a plurality of same blocks, such as a1-a3-a2-a1-a5, in addition to such simple re-sequences. Relative to the number, n, of divided blocks, the number of the former simple re-sequences is n!, while that of the latter complex re-sequences is n^n. That is, for 5 blocks, the number of the former is 5! (5 x 4 x 3 x 2 x 1) of 120, while that of the latter is 5^5 of 3125. Accordingly, the nucleic acid pool as obtained by shuffling a gene according to the present invention thus can cover such an extremely large number of nucleic acid molecules having different base sequences, each of which is different from the base sequence of the original gene.

The kind of the gene to be shuffled is not specifically defined. Employable herein is any and every gene that is composed of polynucleotide chains and contains a coding region necessary for expressing a protein or RNA. The nucleotide unit may contain any molecule of deoxyribonucleotides or ribonucleotides. For the purpose of finding out useful base sequences, preferred are genes coding for proteins, especially enzymes, or control genes for enzymatic functions. Examples of such enzymes include proteases, lipases, cellulases, amylases,

catalases, xylanases, oxidases, dehydrogenases, oxygenases and reductases.

The kind of the gene to which the present invention is directed is not specifically defined but shall be such that, when it is introduced into a suitable host, the host can produce the genetic product through expression of the gene. As examples, referred to are genes as cloned from living organisms, artificially synthesized genes, and even genes as cloned from living organisms and artificially mutated. For the genes derived from living organisms, employable are prokaryotes with definite enzyme producibility. As examples of such prokaryotes, mentioned are bacillus bacteria. One example of the genes derived from such bacteria is a protease API21 gene derived from *Bacillus* NKS-21 (FERM BP-93-1) (Japanese Patent Application Laid-Open No. 5-91876, Sequence Number 1).

Method for Producing Nucleic Acid Pool

The present invention also provides a method for producing a nucleic acid pool comprising two or more different nucleic acids each having a base sequence that is different from the base sequence of the original, non-divided gene, which comprises dividing all or a part of one or more genes into three or more oligonucleotide blocks, followed by ligating all or a part of these blocks into sequences that are different from the sequence or sequences of the original, non-divided gene or genes.

The division of a gene into blocks can be conducted by any desired method that satisfies the above-mentioned conditions

necessary to the nucleic acid pool. For example, the division of a gene can be conducted by the use of restriction enzymes. For this, any desired restriction enzymes can be used, including, for example, EcoRI, HindIII, BamHI, PstI, KpnI, XbaI, SmaI, SacI, ClaI, AluI, HaeIII and RsaI. If a gene having a known base sequence is shuffled, each block of the gene can be obtained through synthesis in accordance with the above-mentioned conditions. To re-sequence these blocks, they are blended and ligated, for example, using a ligase.

Now, preferred methods for producing a single-stranded nucleic acid pool and a double-stranded nucleic acid pool are described in detail hereinunder.

(1) Method for Producing Single-Stranded Nucleic Acid Pool:

One preferred method of producing a single-stranded nucleic acid pool of the present invention is to ligate plural blocks each with a ribonucleotide at the 3'-terminal, using an RNA ligase. This method comprises the following steps a) to c):

a) a step of preparing 3 or more blocks of single-stranded oligonucleotides having base sequences that correspond to all or a part of one or more genes through division of one or more genes or through synthesis of oligonucleotide chains having said base sequences;

b) a step of adding a ribonucleotide to its 3'-terminal of each with no ribonucleotide at the 3'-terminal of the oligonucleotide chain blocks as obtained in the previous step a), while adding a phosphoryl group to its 5'-terminal of each thereof with no phosphoryl group at the 5'-terminal; and

c) a step of ligating in any desired sequence the oligonucleotide chain blocks as obtained in the previous step b), by reacting the 3'-terminal ribonucleotide of one block with the 5'-terminal phosphoryl group of another block.

In Fig. 1, schematically illustrated are the above-mentioned steps for shuffling one gene. In this embodiment illustrated, one gene is divided into four blocks (a1, a2, a3, a4). To simplify the explanation on these steps, the base and the nucleic acids are represented only by the corresponding base sequences. A, G, C and T are nucleotide units comprising the corresponding bases. rG means GMP; and (P) and (OH) mean the phosphoryl group and the hydroxyl group, respectively, existing at the terminals of each nucleotide chain.

Step a)

In the step a), the division of the gene can be effected, using restriction enzymes. After the division, the divided blocks are denatured under heat or with an alkaline or the like into single-stranded oligonucleotide. Where the sequence of the gene is known, single-stranded oligonucleotides are synthesized using ordinary devices and according to ordinary methods.

Step b)

Where the blocks as obtained in the step a) each have a 3'-terminal deoxyribonucleotide, a ribonucleotide is added to the 3'-terminal (step b). This addition can be effected by reacting a terminal deoxynucleotidyl transferase on each said

block in the presence of a nucleoside triphosphate (ATP, GTP, CTP, UTP). The ribonucleotide thus added (AMP, GMP, CMP, UMP) includes the base corresponding to the nucleoside triphosphate used (A, G, C, U). Accordingly, selecting the nucleoside triphosphate to be used, each block may have a desired 3'-terminal ribonucleotide. In the embodiment as illustrated in Fig. 1, GMP (this is represented by rG underlined in Fig. 1) is added to the block mixture. However, if the blocks are separately obtained, for example, by separately synthesizing these, different ribonucleotides can be added to these. The nucleoside triphosphate is used in an amount of from 2 to 10 times or so, by mol, relative to mol of each block. The reaction temperature may be from 30 to 40ºC or so; and the reaction time may be from 30 minutes to 2 hours or so.

In the step b), a phosphoryl group is added to the 5'-terminal of each block. This addition can be effected, using a polynucleotide kinase in the presence of ATP. ATP is used in an amount of from 2 to 10 times or so, by mol, relative to mol of each block. The reaction temperature may be from 30 to 40ºC or so; and the reaction time may be from 10 minutes to 1 hour or so. The pH is most suitably from 7 to 9 or so.

Step c)

The ligation of oligonucleotide chain blocks in the step c) can be effected by reacting an RNA ligase on the mixture of blocks thus obtained in the previous step, in the presence of ATP and divalent metal ions (Japanese Patent Application Laid-Open No. 5-292967). Useful divalent ions are magnesium ions

and manganese ions, of which preferred are magnesium ions. As the ligase, employable is an RNA ligase. The RNA ligase is an enzyme that catalyzes the ligation of a 5'-phosphoryl terminated polynucleotide and a 3'-hydroxyl terminated polynucleotide. The substrate for such an RNA ligase is naturally an RNA, but the enzyme can effectively catalyze the ligation of a 5'-phosphoryl terminated polydeoxyribonucleotide and a polydeoxyribonucleotide having a ribonucleotide only at its 3'-terminal. Preferably used herein is a T4 RNA ligase. The reaction is conducted generally in a buffer, at a pH of from 7 to 9 and at a temperature of from 10 to 40ºC over a period of from 30 to 180 minutes. For example, the oligonucleotides may be reacted in a solution comprising 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 0.1 mM ATP, 10 mg/liter BSA, 1 mM hexaammine cobalt chloride (HCC) and 25 % polyethylene glycol 6000, at 25ºC for 60 minutes or longer.

Controlling of Reading Frame

If the blocks as prepared in the step a) are not n-times (n: integer) the codon units, or if nucleotides are inserted into or deleted from blocks in the step b), the amino acid sequences to be encoded by the blocks vary, depending on the shuffled sites of the blocks. In some cases, however, it is often desirable that the shuffling does not result in the change in the amino acid sequence to be encoded by each block, as so mentioned hereinabove. For this purpose, a modified method as schematically illustrated in Fig. 2 will be effective.

In the modified method illustrated, blocks each having a base sequence of from the (3p+1)th to the (3q+2)th, as counted from the starting point of the reading frame on a gene (where p and q are integers to be independently determined for each block, provided that p ≠ q), are prepared in the first step (step a'), and a ribonucleotide that corresponds to the (3q+3)th base, as counted in the same manner as above, or a ribonucleotide corresponding to a base that does not change the amino acid to be encoded at said site is added to each said block at its 3'-terminal in the step b')

In the embodiment illustrated in Fig. 2, block a1 (p = 0; q = 2), a2 (p = 3; g = 5), a3 (p = 6; q = 9) and a4 are prepared from the gene to be shuffled in the step a') (if desired, these may be divided and isolated). To the block a4, a ribonucleotide is not added at its 3'-terminal for the reasons mentioned below.

Next, GMP (this is represented by rG underlines in Fig. 2) is added to a1 and a2, while AMP (this is represented by rA underlined in Fig. 2) is added to a3, in the next step b') The addition of such ribonucleotides can be effected in the same manner as in the above-mentioned step b), using a nucleoside triphosphate and a terminal deoxynucleotidyl transferase (TDT). Alternatively, employable is a method of preparing ribonucleotide-terminated blocks only. As a result of this step, the amino acid sequences to be encoded by these blocks shall be the same as those on the original gene.

After this, the blocks are phosphorylated with a polynucleotide kinase (PNK) in the same manner as in the above-

mentioned step b) (refer to the latter half of the step b)). Next, in the step c'), these blocks are re-sequenced and ligated in the same manner as in the above-mentioned step c).

Specific Determination of Terminal Sequence

As in Fig. 2, some blocks (a4 in this embodiment) may not be processed with a ribonucleotide at the 3'-terminal. If the blocks not processed so are treated with an RNA ligase, any other block could no more be ligated to these blocks at the 3'-terminal. As a result, all the nucleic acids in the nucleic acid pool obtained shall have substantially any of these blocks at the 3'-terminal. In the same manner, if some particular blocks are not phosphorylated at the 5'-terminal, all the nucleic acids in the nucleic acid pool obtained shall have substantially any of such specific blocks at the 5'-terminal.

Employing this method, it is possible with ease to prepare a nucleic acid pool comprising nucleic acids which have predetermined particular blocks positioned at the terminals while having random re-sequences in the intermediate part. This method is especially advantageous in producing protein mutants having particular amino acid sequences at the terminals or for the purpose of expressing particular control functions.

(2) Method for Producing Double-Stranded Nucleic Acid Pool

The molecules as obtained in the process mentioned in the previous (1) are single-stranded ones, which can be converted into double-stranded ones through genetic treatment thereof to be mentioned below. For this, the block mixture is made to

contain a block having a 5'-phosphoryl group but not having a 3'-ribonucleotide group (this block is referred to as oligonucleotide A; in Fig. 2, a4 corresponds to this block). Accordingly, all the nucleic acids that constitute the nucleic acid pool to be produced shall be substantially terminated by oligonucleotide A at the 3'-terminal, as so mentioned in the last in the previous (2). Next, the nucleic acid blocks are subjected to ordinary DNA-extending reaction, using, as a primer, a decamer (10-mer) or higher oligonucleotide, preferably a heptamer (17-mer) or higher oligonucleotide, that is complementary to oligonucleotide A. For this, employable is any and every enzyme that catalyzes the DNA-extending reaction, such as Taq polymerase, Klenow fragment, DNA polymerase I or the like.

For this purpose, also employable is PCR (polymerase chain reaction). If PCR is employed, an additional oligonucleotide having a 3'-ribonucleotide group but not having a 5'-phosphoryl group (hereinafter referred to as oligonucleotide B) is added to the block mixture, in addition to the above-mentioned oligonucleotide A, during the process of preparing the pool. Accordingly, all the molecules that constitute the pool shall have oligonucleotide A at the 3'-terminal and oligonucleotide B at the 5'-terminal.

After this, the nucleic acid blocks are subjected to PCR, using, as primers, a 10-mer or higher oligonucleotide, preferably a 17-mer or higher oligonucleotide, that is complementary to oligonucleotide A, and a 10-mer or higher oligonucleotide, preferably a 17-mer or higher oligonucleotide,

that is complementary to oligonucleotide B, whereby the nucleic acid blocks are converted into double-stranded ones while being amplified at the same time. Therefore, this process is advantageous for the following operation.

The oligonucleotide A and/or B may be the same as those existing on the original gene, or, if desired, may also be others which the original gene does not have.

Expression of Genetic Information in Nucleic Acid Pool

The resulting double-stranded nucleic acid is blunted, and then ligated to any desired vector, preferably an expression vector, such as pKK223-3, using a DNA ligase. If desired, the polynucleotide A and B positioned at the both terminals of the nucleic acid may be made to have suitable restriction enzyme recognizing sites. In this case, the nucleic acid may be ligated to a suitable vector, using the defined restriction enzymes.

Next, the vector library thus produced in the manner mentioned above is introduced into a suitable host, in which the genetic information is expressed. Thus, the intended genetic product with favorable properties and also the gene coding for it can be obtained. Any and every ordinary host can be used herein. Preferred examples of the host include cells of *E. coli*, bacillus bacteria, yeasts, and lactic acid bacteria.

If desired, *in-vitro* transcription systems and translation systems are also employable herein. In those cases, the genetic information can be expressed even when the nucleic acid is not ligated to a vector.

The "genetic information" as referred to herein indicates the information on a gene which is carried by a DNA or RNA and which is translated into a protein or is transcribed into RNA in a suitable living body by the DNA or RNA for itself or after having been ligated to any other DNA or RNA.

The genetic information that is expected to be expressed according to the method of the present invention is not specifically defined, but includes, for example, those on various genetic products, such as enzymes, antibodies, hormones receptor proteins and ribozymes, and those on various control functions of, for example, operators, promoters and attenuators.

Examples

Now, the present invention is described more concretely hereinunder with reference to the following examples, which, however, are not intended to restrict the scope of the present invention.

Example 1: Production of Single-Stranded Nucleic Acid Pool

A nucleic acid pool was produced in accordance with the process mentioned below, based on the wild-type alkali protease (Japanese Patent Application-Laid Open No. 5-91876) as cloned from a protease API21 (*Bacillus* NKS-21; FERM BP-93-1) having a sequence of Sequence Number 1.

(1) Step a): Preparation of Oligonucleotide Blocks

Using an automatic DNA synthesizer, Model 392 (manufactured by Perkin Elmer Co.), synthesized were the following 5 oligonucleotides.

① Oligo A (Sequence Number 2; this corresponds to the base sequence of from 436th to 455th in Sequence Number 1)

② Oligo 1 (Sequence Number 3; from 457th to 503rd in Sequence Number 1)

③ Oligo 2 (Sequence Number 4; from 505th to 551st in Sequence Number 1)

④ Oligo 3 (Sequence Number 5; from 553rd to 596th in Sequence Number 1)

⑤ Oligo B (Sequence Number 6; from 598th to 618th in Sequence Number 1)

Oligo A, oligo 1 to 3, and oligo B are parts of the protease API21 gene. Their positions are as mentioned above. These oligonucleotides were synthesized in a DM trityl-on condition (that is, while the 5'-hydroxyl group was protected with dimethoxytrityl group), and purified through an OPC column. The reagents used herein were obtained from Perkin Elmer Co.

(2) Step b: Processing of Oligonucleotide Blocks

(2-1) Addition of Ribonucleotide:

500 pmols of oligo A, 1 nmol of UTP and 10 units of terminal deoxynucleotidyl transferase were added to a standard solution comprising:

50 mM Tris-HCl buffer (pH 8.0)

10 mM $MgCl_2$

5 mM DTT (dithiothreitol)

25 % PEG 6000

1 mM HCC (hexaammine cobalt chloride)

10 μg/ml BSA (bovine serum albumin),

to thereby make 10 μl in total. The resulting solution was left at 37ºC for 1 hour.

Oligo 1 and oligo 2 were processed in the same manner as above. Oligo 3 was processed in the same manner as above, but using ATP in place of UTP. These oligonucleotide blocks to which had been added 3'-terminal ribonucleotide through the above-mentioned operation are referred to as oligo Ar, oligo 1r, oligo 2r and oligo 3r.

(2-2) Phosphorylation:

500 pmols of oligo 1r, 1 nmol of ATP and 10 units of polynucleotide kinase were dissolved in the standard solution having the same composition as above to make 10 μl in total. The resulting solution was left at 37ºC for 1 hour. Oligo 2r, oligo 3r and oligo B were processed in the same manner as above. These polynucleotides thus formed are referred to as oligo 1pr, oligo 2pr, oligo 3pr and oligo Bp.

(3) Step c): Ligation of Oligonucleotide Blocks:

500 pmols of oligo Ar, 500 pmols of oligo 1pr, 500 pmols of oligo 2pr, 500 pmols of oligo 3pr, 500 pmols of oligo Bp, that had been prepared in the previous step, and also 1 nmol of ATP and 50 units of T4 RNA ligase were dissolved in the standard solution having the same composition as above to make 10 μl in total. These were thus reacted at 25ºC for 4 hours.

Subsequently, the reaction mixture was subjected to polyacrylamide gel electrophoresis, through which were collected fragments of about 180 bp from the gel.

As a result, obtained was a single-stranded nucleic acid pool in which oligo 1pr, oligo 2pr and oligo 3pr were ligated in random sequences between oligo Ar and oligo Bp.

Example 2: Production of Double-Stranded Nucleic Acid Pool

Oligo B' (its sequence is represented by Sequence Number 7) which is complementary to oligo B was synthesized in the same manner as in Example 1.

The DNAs' constituting the single-stranded nucleic acid pool as obtained in Example 1 were all or, that is, without being separated into the individual DNAs, mixed with 10 pmols of oligo B', and added to tris-HCl buffer containing $MgCl_2$ and DTT (dithiothreitol) to make 20 μl in total. The resulting mixture was finally comprised of 10 mM tris-HCl buffer (pH 7.5), 7 mM $MgCl_2$, and 0.1 mM of DTT. This was heated at 75ºC for 5 minutes, and then cooled to 30ºC. Next, 1 unit of Klenow fragment was added thereto, and kept at 37ºC for 2 hours, whereby the single-stranded nucleic acids were converted into double-stranded ones.

As a result, obtained was a double-stranded nucleic acid pool in which oligo 1pr, oligo 2pr and oligo 3pr were ligated in random sequences between oligo Ar and oligo Bp.

Example 3: Transformation of *E. coli* with Nucleic Acid Pool

A plasmid pSDT812 (Japanese Patent Application Laid-Open No. 1-141596), which has been prepared by inserting a gene of a wild-type alkali protease as cloned from *Bacillus* NKS-21 into a plasmid pHSG396 at its ClaI-cleaving site, was digested with a restriction enzyme ClaI, and then blunted with a commercially-available blunting reagent (Blunting Kit, manufactured by Takara Shuzo Co.). This was mixed with a plasmid pHY300PLK (manufactured by Yakult Honsha Co.), which had been digested with restriction enzymes EcoRI and HindIII, then blunted in the same manner as above and processed with an alkali phosphatase, and these were ligated using a commercially-available ligation kit (manufactured by Takara Shuzo Co.).

Using the resulting DNA, cells of *E. coli* JM105 were transformed, and tetracycline-resistant transformants were selected. From these transformants, the plasmid DNA was extracted, purified and analyzed. Thus was obtained a plasmid pHY812 (Fig. 3), in which pHY300PLK was ligated to the wild-type alkali protease.

Next, pHY812 formed in the above was digested with restriction enzymes HindIII and SphI, and then processed with BAP (a). On the other hand, the DNAs constituting the double-stranded nucleic acid pool as formed in Example 3 were digested all at a time with restriction enzymes HindIII and SphI (b). (a) and (b) were ligated in the same manner as above, using the ligation kit. With the resulting DNA, cells of *E. coli* JM105

were transformed. The resulting transformants were screened on an L-plate containing ampicillin.

Example 4

From the colonies that had been selected in Example 3, a plasmid DNA was prepared, which was then digested with HindIII and SphI to check as to whether or not it gave a fragment of about 160 bp after the digestion. The base sequences of 95 clones that had given the fragment having the intended length were analyzed, which verified the shuffling of the gene blocks corresponding to oligo 1, oligo 2 and oligo 3. Table 1 shows different types of shuffling, and the number of clones with each type.

Table 1

Type of Shuffling	Number of Clones	Type of Shuffling	Number of Clones
111	3	223	3
112	4	231	6
113	2	232	2
121	5	233	3
122	3	311	3
123	6	312	5
131	2	313	3
132	7	321	7
133	3	322	3
211	2	323	4
212	3	331	2
213	5	332	5
221	2	333	1
222	1		

As in the above, it has been confirmed that, if three blocks of one gene are shuffled according to the method of the present invention, obtained are all combinations of clones each containing the same or different three of these blocks.

Example 5

The plasmid DNAs as produced in Example 4 were mixed. Using the resulting DNA mixture, cells of *Bacillus subtilis* UOTO999 were transformed. Tetracycline-resistant transformants

were selected. 300 transformants were replicated on a skim milk-containing medium plate, on which were found clear zones around the colonies of 15 transformants. Accordingly, it is understood that the enzyme which the shuffled gene codes for can be selected depending on its activity. A plasmid DNA was prepared from these 15 transformants that had formed the clear zones, and then sequenced. From the base sequence thus identified, it is understood that the blocks of the plasmid DNA prepared herein were sequenced in the same order as in the wild-type plasmid DNA.

Example 6

From 10 transformants (one forms clear zones, while nine do not) as obtained in Example 5, and also from the host, *Bacillus subtilis* UOTO999 which does not have the plasmid, full-length RNAs were prepared. These were processed with a ribonuclease-free deoxyribonuclease, in order to remove the influence of the plasmid on the hybridization to be effected later on. Next, using oligo B' as the probe, these were subjected to Northern hybridization. As a result, all lanes corresponding to the RNA of the transformants gave detectable bands, but no band was detected on the lanes corresponding to the RNA of the host.

[Advantages of the Invention]

According to the present invention, it is possible to obtain, through simple processes, a nucleic acid pool capable of covering various base sequences which are substantially

apart from the naturally-existing base sequence spaces. Therefore, it is possible to obtain excellent genetic products, such as proteins and enzymes, which could not be obtained in conventional methods and which were not examined by organisms in the past. In addition, according to the method of the present invention for producing a nucleic acid pool, it is possible to obtain a mixture of nucleic acids while optionally shuffling the constitutive blocks at random in the intermediate parts but fixing the terminal sequences to be predetermined, desired ones, and it is also possible to shuffle the constitutive blocks without changing the amino acid sequence which each block codes for. Therefore, as compared with a method of producing a completely-randomized nucleic acid pool, there is a high possibility that useful genetic products can be produced according to the method of the present invention.

Sequence Listing

Sequence Number: 1

Length of Sequence: 1122

Type of Sequence: Nucleic Acid

Number of Strands: Double-stranded

Topology: Linear

Kind of Sequence: Genomic DNA

Source: *Bacillus* NKS-21 (FERM BP-93-1)

Characteristics of Sequence:

Code Indicating Characteristics: Sig Peptide

Existing Site: 1 . . . 93

Method of Determining Characteristics: S

Code Indicating Characteristics: Mat Peptide

Existing Site: 104 . . . 1112

Method of Determining Characteristic: S

Sequence:

```
ATG AAT CTT CAA AAA ATA GCC TCA GCG TTG AAG GTT AAG CAA TCG GCA48
Met Asn Leu Gln Lys Ile Ala Ser Ala Leu Lys Val Lys Gln Ser Ala
        -100                -95                 -90
TTG GTC AGC AGT TTA ACT ATT TTG TTT CTA ATC ATG CTA GTA GGT ACG96
Leu Val Ser Ser Leu Thr Ile Leu Phe Leu Ile Met Leu Val Gly Thr
    -85                 -80                 -75
ACT AGT GCA AAT GGT GCG AAG CAA GAG TAC TTA ATT GGT TTC AAC TCA 144
Thr Ser Ala Asn Gly Ala Lys Gln Glu Tyr Leu Ile Gly Phe Asn Ser
-70                 -65                 -60                 -55
GAC AAG GCA AAA GGA CTT ATC CAA AAT GCA GGT GGA GAA ATT CAT CAT 192
Asp Lys Ala Lys Gly Leu Ile Gln Asn Ala Gly Gly Glu Ile His His
                -50                 -45                 -40
GAA TAT ACA GAG TTT CCA GTT ATC TAT GCA GAG CTT CCA GAA GCA GCG 240
Glu Tyr Thr Glu Phe Pro Val Ile Tyr Ala Glu Leu Pro Glu Ala Ala
            -35                 -30                 -25
GTA AGT GGA TTG AAA AAT AAT CCT CAT ATT GAT TTT ATT GAG GAA AAC 288
Val Ser Gly Leu Lys Asn Asn Pro His Ile Asp Phe Ile Glu Glu Asn
        -20                 -15                 -10
GAA GAA GTT GAA ATT GCA CAG ACT GTT CCT TGG GGA ATC CCT TAT ATT 336
Glu Glu Val Glu Ile Ala Gln Thr Val Pro Trp Gly Ile Pro Tyr Ile
     -5                  1                   5                  10
```

```
TAC TCG GAT GTT GTT CAT CGT CAA GGT TAC TTT GGG AAC GGA GTA AAA 384
Tyr Ser Asp Val Val His Arg Gln Gly Tyr Phe Gly Asn Gly Val Lys
                 15                  20                  25
GTA GCA GTA CTT GAT ACA GGA GTG GCT CCT CAT CCT GAT TTA CAT ATT 432
Val Ala Val Leu Asp Thr Gly Val Ala Pro His Pro Asp Leu His Ile
             30                  35                  40
AGA GGA GGA GTA AGC TTT ATC TCT ACA GAA AAC ACT TAT GTG GAT TAT 480
Arg Gly Gly Val Ser Phe Ile Ser Thr Glu Asn Thr Tyr Val Asp Tyr
         45                  50                  55
AAT GGT CAC GGT ACT CAC GTA GCT GGT ACT GTA GCT GCC CTA AAC AAT 528
Asn Gly His Gly Thr His Val Ala Gly Thr Val Ala Ala Leu Asn Asn
     60                  65                  70
TCA TAT GGC GTA TTG GGA GTG GCT CCT GGA GCT GAA CTA TAT GCT GTT 576
Ser Tyr Gly Val Leu Gly Val Ala Pro Gly Ala Glu Leu Tyr Ala Val
 75                  80                  85                  90
AAA GTT CTT GAT CGT AAC GGA AGC GGT TCG CAT GCA TCC ATT GCT CAA 624
Lys Val Leu Asp Arg Asn Gly Ser Gly Ser His Ala Ser Ile Ala Gln
                 95                 100                 105
GGA ATT GAA TGG GCG ATG AAT AAT GGG ATG GAT ATT GCC AAC ATG AGT 672
Gly Ile Glu Trp Ala Met Asn Asn Gly Met Asp Ile Ala Asn Met Ser
            110                 115                 120
TTA GGA AGT CCT TCT GGG TCT ACA ACC CTG CAA TTA GCA GCA GAC CGC 720
Leu Gly Ser Pro Ser Gly Ser Thr Thr Leu Gln Leu Ala Ala Asp Arg
        125                 130                 135
GCT AGG AAT GCA GGT GTC TTA TTA ATT GGG GCG GCT GGA AAC TCA GGA 768
Ala Arg Asn Ala Gly Val Leu Leu Ile Gly Ala Ala Gly Asn Ser Gly
    140                 145                 150
CAA CAA GGC GGC TCG AAT AAC ATG GGC TAC CCA GCG CGC TAT GCA TCT 816
Gln Gln Gly Gly Ser Asn Asn Met Gly Tyr Pro Ala Arg Tyr Ala Ser
155                 160                 165                 170
GTC ATG GCT GTT GGA GCG GTG GAC CAA AAT GGA AAT AGA GCG AAC TTT 864
Val Met Ala Val Gly Ala Val Asp Gln Asn Gly Asn Arg Ala Asn Phe
                175                 180                 185
TCA AGC TAT GGA TCA GAA CTT GAG ATT ATG GCG CCT GGT GTC AAT ATT 912
Ser Ser Tyr Gly Ser Glu Leu Glu Ile Met Ala Pro Gly Val Asn Ile
            190                 195                 200
AAC AGT ACG TAT TTA AAT AAC GGA TAT CGC AGT TTA AAT GGT ACG TCA 960
Asn Ser Thr Tyr Leu Asn Asn Gly Tyr Arg Ser Leu Asn Gly Thr Ser
        205                 210                 215
ATG GCA TCT CCA CAT GTT GCT GGG GTA GCT GCA TTA GTT AAA CAA AAA1008
Met Ala Ser Pro His Val Ala Gly Val Ala Ala Leu Val Lys Gln Lys
    220                 225                 230
CAC CCT CAC TTA ACG GCG GCA CAA ATT CGT AAT CGT ATG AAT CAA ACA1056
His Pro His Leu Thr Ala Ala Gln Ile Arg Asn Arg Met Asn Gln Thr
235                 240                 245                 250
GCA ATT CCG CTT GGT AAC AGC ACG TAT TAT GGA AAT GGC TTA GTG GAT1104
Ala Ile Pro Leu Gly Asn Ser Thr Tyr Tyr Gly Asn Gly Leu Val Asp
                255                 260                 265
GCT GAG TAT GCG GCT CAA                                     1122
Ala Glu Tyr Ala Ala Gln
            270     272
```

Sequence Number: 2

Length of Sequence: 20

Type of Sequence: Nucleic Acid

Number of Strand: Single-stranded

Topology: Linear

Kind of Sequence: Other Nucleic Acid, Synthetic DNA

Sequence:

GGAGGAGTAA GCTTTATCTC

Sequence Number: 3

Length of Sequence: 47

Type of Sequence: Nucleic Acid

Number of Strand: Single-stranded

Topology: Linear

Kind of Sequence: Other Nucleic Acid, Synthetic DNA

Sequence:

ACAGAAAACA CTTATGTGGA TTATAATGGT CACGGTACTC ACGTAGC

Sequence Number: 4

Length of Sequence: 47

Type of Sequence: Nucleic Acid

Number of Strand: Single-stranded

Topology: Linear

Kind of Sequence: Other Nucleic Acid, Synthetic DNA

Sequence:

GGTACTGTA GCTGCCCTAA ACAATTCATA TGGCGTATTG GGAGTGGC

Sequence Number: 5

Length of Sequence: 44

Type of Sequence: Nucleic Acid

Number of Strand: Single-stranded

Topology: Linear

Kind of Sequence: Other Nucleic Acid, Synthetic DNA

Sequence:

CCTGGAGCTG AACTATATGC TGTTAAAGTT CTTGATCGTA ACGG

Sequence Number: 6

Length of Sequence: 21

Type of Sequence: Nucleic Acid

Number of Strand: Single-stranded

Topology: Linear

Kind of Sequence: Other Nucleic Acid, Synthetic DNA

Sequence:

AGCGGTTCGC ATGCATCCAT T

Sequence Number: 7

Length of Sequence: 21

Type of Sequence: Nucleic Acid

Number of Strand: Single-stranded

Topology: Linear

Kind of Sequence: Other Nucleic Acid, Synthetic DNA

Sequence:

AATGGATGCA TGCGAACCGCT

Brief Description of the Drawings

Fig. 1 is a schematic view showing one embodiment of the method for producing a nucleic acid pool of the present invention.

Fig. 2 is a schematic view showing another embodiment of the method for producing a nucleic acid pool of the present invention.

Fig. 3 is a restriction enzyme cleavage map of plasmid pHY812, in which the alkali protease gene derived from Bacillus NKS-21 has been ligated to plasmid pHY300PLK.

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM
(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 14 , lines 14-15 to page , lines

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

National Institute of Bioscience and Human-Technology, Agency of Industrial Science and Technology, Ministry of International Trade and Industry

Address of depositary institution *(including postal code and country)*

1-3 Higashi 1-chome, Tsukuba-shi, Ibaraki-ken, Japan

Date of deposit	Accession Number
7 May 1985	FERM BP-93-1

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional ☐

Until the publication of the mention of grant of a European patent or, where applicable, for twenty years from the date of filing if the application has been refused, withdrawn or deemed withdrawn, a sample of the deposited microorganism is only to be provided to an independent expert nominated by the person requesting the sample (cf. Rule 28(4) EPC). And as far as Australia is concerned, the expert option is likewise requested, reference being had to Regulation 3.25 of Australia Statutory Rules 1991 No 71. Also, for Canada we request that only an independent expert nominated by the Commissioner is authorized to have access to a sample of the microorganism deposited.

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only

☒ This sheet was received with the international application

Authorized officer Anne-Grethe Warrer-Madsen
Head Clerk

A. G. Warrer-Madsen

For International Bureau use only

☐ This sheet was received by the International Bureau on:

Authorized officer

Form PCT/RO/134 (July 1992)

Claims

1. A nucleic acid pool comprising two or more different nucleic acids, which is constructed by dividing all or a part of one or more genes into 3 or more blocks followed by ligating all or a part of these blocks into sequences that are different from the sequence or sequences of the original, non-divided gene or genes.

2. The nucleic acid pool as claimed in claim 1, containing 10 or more different nucleic acids.

3. The nucleic acid pool as claimed in claim 1 or 2, containing all the nucleic acids with different sequences as constructed by re-sequencing a number, n, of said blocks (where n is the number of the different blocks as formed by the division).

4. The nucleic acid pool as claimed in any one of claims 1 to 3, wherein the gene is a gene coding for a protein, and the amino acid sequence as encoded by each block is the same as the amino acid sequence as encoded by the corresponding part on the original gene.

5. The nucleic acid pool as claimed in any one of claims 1 to 4, wherein the gene is a gene coding for an enzymatic function or a control gene for it.

6. The nucleic acid pool as claimed in claim 5, wherein the gene is a gene coding for any one of proteases, lipases, cellulases, amylases, catalases, xylanases, oxidases, dehydrogenases, oxygenases and reductases.

7. The nucleic acid pool as claimed in any one of claims 1 to 6, wherein the gene is one derived from prokaryotes.

8. The nucleic acid pool as claimed in claim 7, wherein the gene is one derived from bacillus bacteria.

9. The nucleic acid pool as claimed in claim 8, wherein the gene is a protease API21 gene.

10. The nucleic acid pool as claimed in any one of claims 1 to 9, wherein each block is an oligonucleotide.

11. The nucleic acid pool as claimed in claim 10, wherein the nucleic acid is a single-stranded polynucleotide.

12. The nucleic acid pool as claimed in claim 10, wherein the nucleic acid is a double-stranded polynucleotide.

13. A method for producing a nucleic acid pool comprising two or more different nucleic acids, which comprises dividing all or a part of one or more genes into three or more oligonucleotide blocks or synthesizing oligonucleotides corresponding to said blocks, followed by ligating all or a part of these blocks into sequences that are different from those on genes.

14. The method for producing a nucleic acid pool as claimed in claim 13, which comprises the following steps a) to c):

a) a step of preparing 3 or more blocks of single-stranded oligonucleotides having base sequences that correspond to all or a part of one or more genes through division of one or more genes or through synthesis of oligonucleotide chains having said base sequences;

b) a step of adding a ribonucleotide to its 3'-terminal of each with a deoxyribonucleotide at the 3'-terminal of the oligonucleotide chain blocks as obtained in the previous step

a), while adding a phosphoryl group to its 5'-terminal of each thereof with a hydroxyl group at the 5'-terminal; and

c) a step of ligating in any desired sequence the oligonucleotide chain blocks as obtained in the previous step b), by reacting the 3'-terminal ribonucleotide of one block with the 5'-terminal phosphoryl group of another block.

15. The method for producing a nucleic acid pool as claimed in claim 14, wherein the number of the blocks to be prepared in the step a) is 3 or more.

16. The method for producing a nucleic acid pool as claimed in claim 14 or 15, wherein at least one block is left to still have its 5'-terminal hydroxyl group in the step b) to thereby selectively obtain a nucleic acid or nucleic acids having said block at the 5'-terminal.

17. The method for producing a nucleic acid pool as claimed in any one of claims 14 to 16, wherein at least one block is left to still have its 3'-terminal deoxyribonucleotide in the step b) to thereby selectively obtain a nucleic acid or nucleic acids having said block at the 3'-terminal.

18. The method for producing a nucleic acid pool as claimed in any one of claim 14 to 17, wherein the blocks are prepared in such a manner that the amino acid sequence as encoded by each block is the same as the amino acid sequence as encoded by the corresponding part on the original gene.

19. The method for producing a nucleic acid pool as claimed in claim 18, wherein blocks each having a base sequence of from the (3p+1)th to the (3q+2)th, as counted from the starting point of the reading frame on a gene (where p and q

are integers to be independently determined for each block, provided that p , q), are prepared in the step a), and a ribonucleotide that corresponds to the (3q+3)th base, as counted in the same manner as above, or corresponds to a base that does not change the amino acid to be encoded at said site is added to each said block at its 3'-terminal in the step b).

20. The method for producing a nucleic acid pool as claimed in any one of claims 14 to 19, wherein the ligation of the step c) is conducted, using an RNA ligase in the presence of adenosine triphosphate and divalent metal ions.

21. A method for producing a double-stranded nucleic acid pool, which comprises converting the single-stranded nucleic acids as obtained in any one of claims 14 to 20 into double-stranded ones through polymerase reaction.

22. A genetic product to be obtained by expressing the genetic information that exists in the nucleic acid pool as set forth in any one of claims 1 to 12.

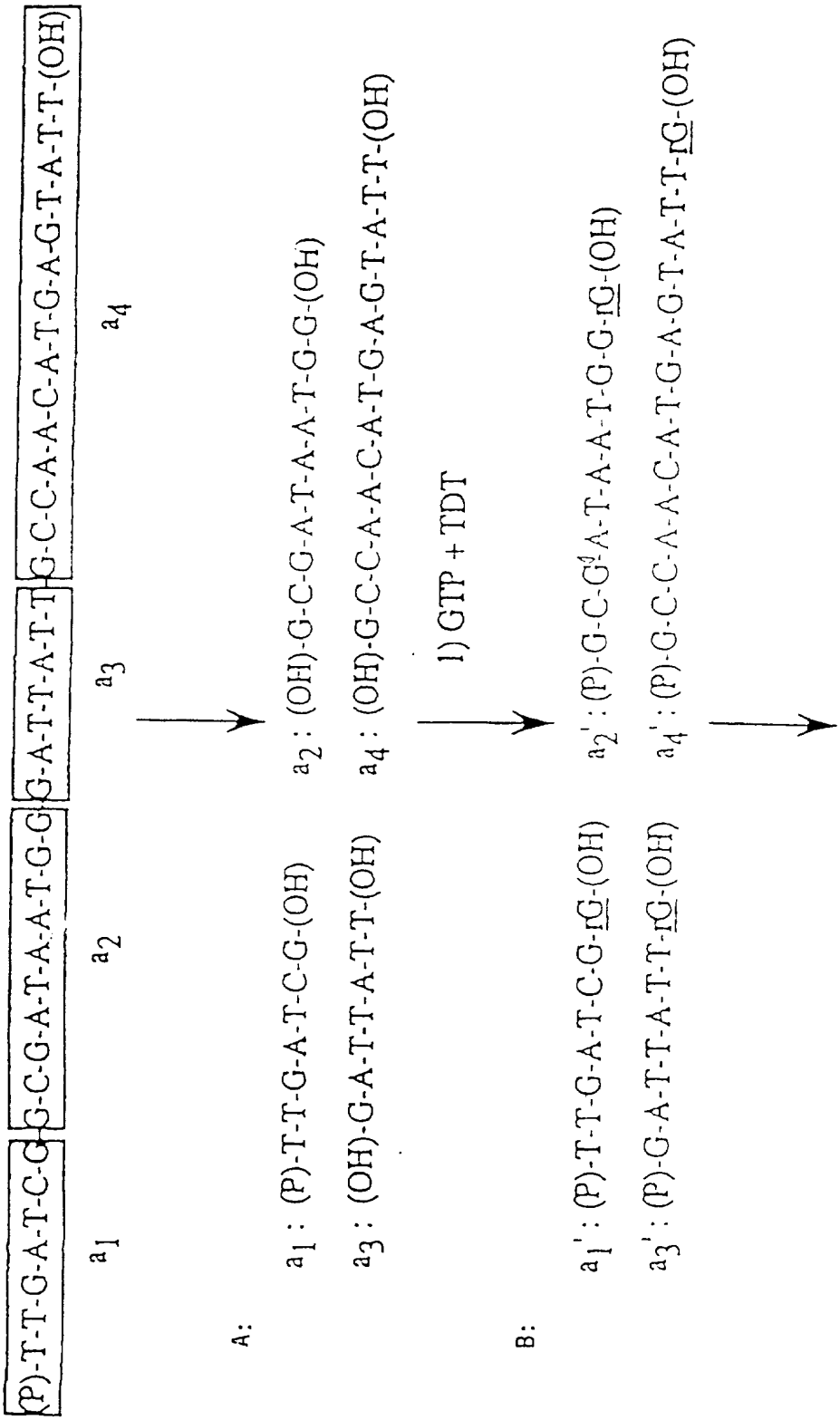
(P)-T-T-G-A-T-C-G
G-C-G-A-T-A-A-T-G-G
G-A-T-T-A-T-T
G-C-C-A-A-C-A-T-G-A-G-T-A-T-T-(OH)
a1
a2
a3
a4
A:
a1 : (P)-T-T-G-A-T-C-G-(OH)
a3 : (OH)-G-A-T-T-A-T-T-(OH)
a2 : (OH)-G-C-C-A-A-C-A-T-G-A-G-T-A-T-T-(OH)
a4 : (OH)-G-C-C-A-A-C-A-T-G-A-G-T-A-T-T-(OH)
1) GTP + TDT
B:
a1' : (P)-T-T-G-A-T-C-G-rG-(OH)
a3' : (P)-G-A-T-T-A-T-T-rG-(OH)
a2' : (P)-G-C-C-A-A-C-A-T-G-A-G-T-A-T-T-rG-(OH)
a4' : (P)-G-C-C-A-A-C-A-T-G-A-G-T-A-T-T-rG-(OH)
C:
a1', a2', a3', a4'

Fig. 1

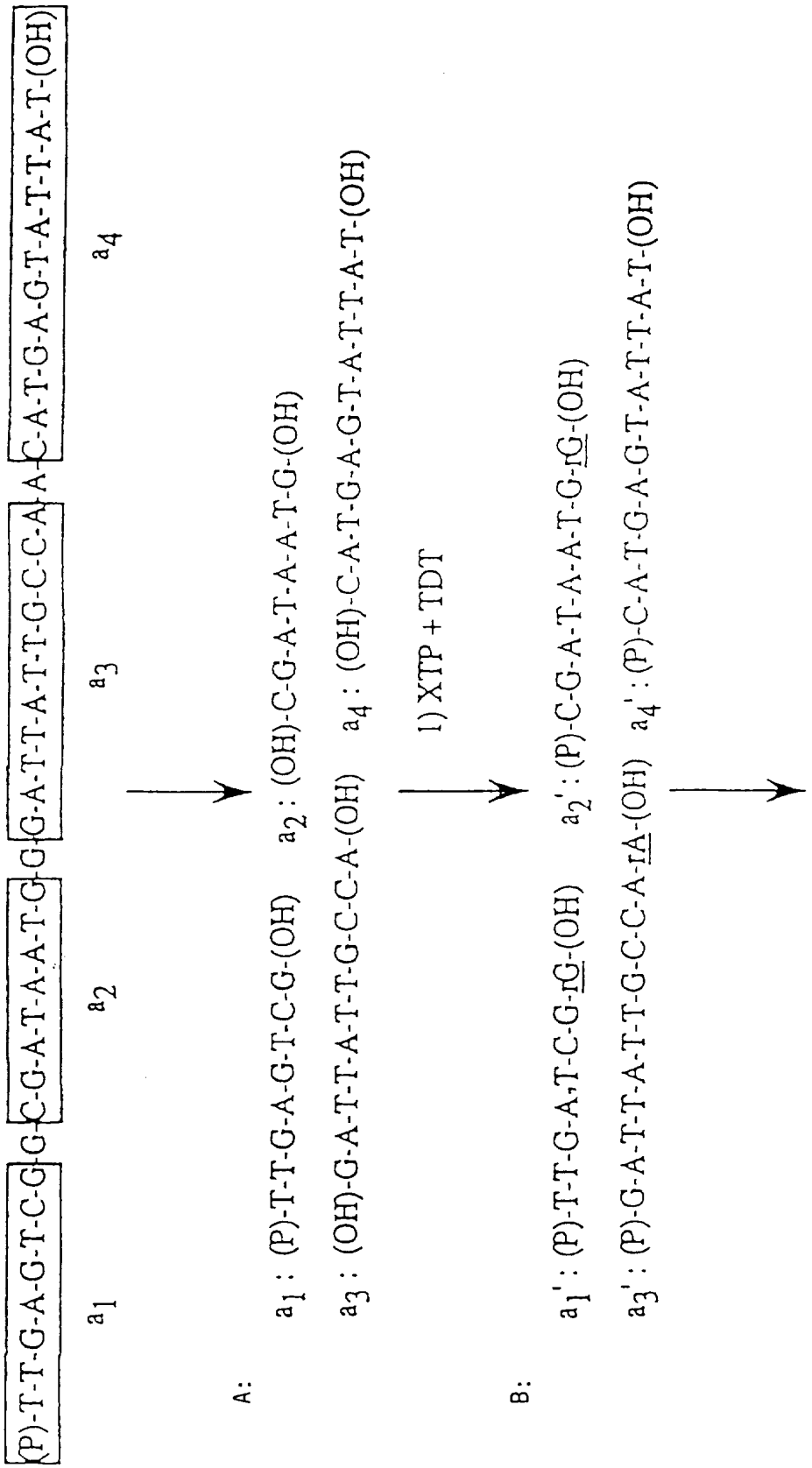
(P)-T-T-G-A-G-T-C-G-C-G-A-T-A-A-T-G-G-A-T-T-A-T-T-G-C-C-A-C-A-T-G-A-G-T-A-T-T-A-T-(OH)
a1
a2
a3
a4
A:
a1 : (P)-T-T-G-A-G-T-C-G-(OH)
a2 : (OH)-C-G-A-T-A-A-T-G-(OH)
a3 : (OH)-G-A-T-T-A-T-T-G-C-C-A-(OH)
a4 : (OH)-C-A-T-G-A-G-T-A-T-T-A-T-(OH)
1) XTP + TDT
B:
a1' : (P)-T-T-G-A,T-C-G-rG-(OH)
a2' : (P)-C-G-A-T-A-A-T-G-rG-(OH)
a3' : (P)-G-A-T-T-A-T-T-G-C-C-A-rA-(OH)
a4' : (P)-C-A-T-G-A-G-T-A-T-T-A-T-(OH)
C:
a4'

Fig. 2

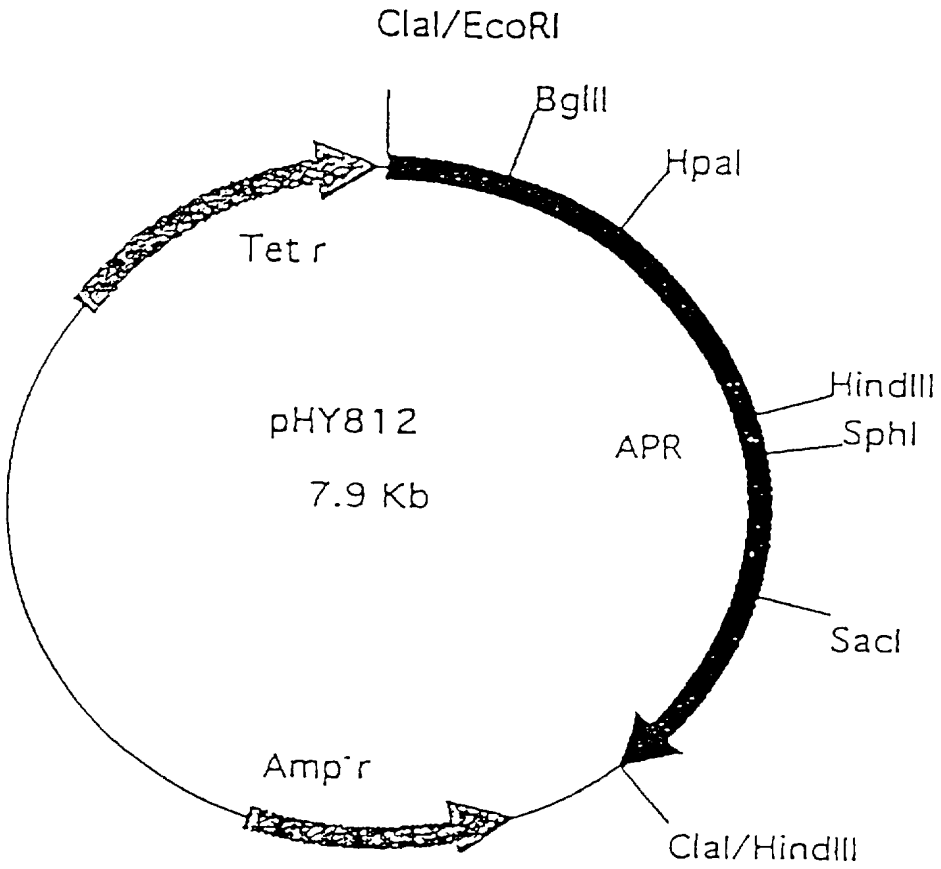
ClaI/EcoRI
BglII
HpaI
Tet r
HindIII
SphI
pHY812
7.9 Kb
APR
SacI
Amp r
ClaI/HindIII

Fig. 3

INTERNATIONAL SEARCH REPORT

Inte...ational application No.
PCT/DK 97/00316

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: C12N 15/10, C12Q 1/68
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: C12N, C07K, C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, MEDLINE, BIOSIS, DBA, CA, SCISEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
D,X	WO 9522625 A1 (AFFYMAX TECHNOLOGIES N.V.), 24 August 1995 (24.08.95), see especially page 44, line 34 - page 45, line 15 --	1-22
X	WO 9517413 A1 (EVOTEC BIOSYSTEMS GMBH), 29 June 1995 (29.06.95), see especially page 8, line 3-7; page 9, line 15-16; page 16, line 7-11; and claims --	1-22
A	WO 9218645 A1 (DIAGEN INSTITUT FüR MOLEKULARBIOLOGISCHE DIAGNOSTIK GMBH), 29 October 1992 (29.10.92) --	1-22

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 November 1997	19 -11- 1997
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Patrick Andersson Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

01/10/97

International application No.
PCT/DK 97/00316

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9522625 A1	24/08/95	AU 2971495 A	04/09/95
		CA 2182393 A	24/08/95
		CN 1145641 A	19/03/97
		EP 0752008 A	08/01/97
		US 5605793 A	25/02/97
WO 9517413 A1	29/06/95	DE 4343591 A	22/06/95
WO 9218645 A1	29/10/92	DE 4112440 C	22/10/92
		EP 0583265 A	23/02/94
		JP 6506355 T	21/07/94